# In-Depth Benchmark Report & Legal Complaint

Llama 3.2 3B Chat on Microsoft Surface Pro 11 (Snapdragon X Elite)

## 1. Introduction & Motivation

I have conducted independent testing of Qualcomm's Snapdragon X Elite processor in the Microsoft Surface Pro 11, specifically assessing its Neural Processing Unit (NPU) performance in local large language model (LLM) inference. The NPU is marketed as delivering high-speed on-device AI acceleration for models such as ChatGPT-like assistants. However, based on reproducible benchmarking, I found the NPU to be significantly slower than the CPU for this purpose. This contradicts Qualcomm's and Microsoft's marketing claims and raises concerns under Australian Consumer Law (ACL).

## 2. Hardware and Software Setup

### Component Specification

| | |
|---|---|
| Device | Microsoft Surface Pro 11 (11th Gen) |
| Processor | Qualcomm Snapdragon X Elite (X1E) |
| RAM | 16 GB LPDDR5 |
| Storage | 512 GB NVMe SSD |
| Operating System | Windows 11 ARM64 |
| Software | AnythingLLM v1.8.4 |
| Model Tested | Llama 3.2 3B Chat |
| Context Window | 8,192 tokens |

## 3. Benchmark Test Procedure

Two benchmark runs were performed using the same prompt:
"List benefits of renewable energy."

Measured metrics:
- Total output time
- Token count
- Tokens per second (throughput)

Comparison:

# In-Depth Benchmark Report & Legal Complaint

Llama 3.2 3B Chat on Microsoft Surface Pro 11 (Snapdragon X Elite)

- NPU-accelerated inference
- CPU-only inference

## 4. Results

a) NPU Inference
Time Taken: 123.333 sec
Throughput: 2.64 tokens/sec
Output: List of 10 renewable energy benefits (complete, but very slow)

b) CPU Inference
Time Taken: 10.277 sec
Throughput: 25.88 tokens/sec
Output: List of 8 renewable energy benefits (coherent and informative)

## 5. Performance Summary

CPU inference was over 12x faster than NPU.
NPU did not produce better output quality to justify the extreme slowdown.
No practical advantage of NPU in current LLM workflows.
Marketing promises of "AI acceleration" are not realized in actual use cases.

## 6. Technical Analysis of Failure

The Hexagon NPU in Snapdragon X Elite is not yet optimized for transformer models.
Latency from data transfer between CPU and NPU nullifies theoretical gains.
AnythingLLM and ONNX runtimes fail to leverage NPU effectively due to lack of mature software stack.
Qualcomm's documentation does not provide real-world LLM benchmarks for NPU use.

## 7. Legal Complaint under Australian Consumer Law (ACL)

Claim: Misleading or Deceptive Conduct (ACL Section 18)
Microsoft and Qualcomm jointly marketed the Surface Pro 11 with Snapdragon X Elite using phrases such as:

# In-Depth Benchmark Report & Legal Complaint

Llama 3.2 3B Chat on Microsoft Surface Pro 11 (Snapdragon X Elite)

"...run AI models with industry-leading performance and efficiency thanks to the integrated NPU..."
"...designed for the AI era..."
"...blazingly fast local AI experiences..."

These claims are demonstrably false in the context of LLM inference.

Claim: False or Misleading Representations (ACL Section 29)
Under Section 29(1)(g) of the ACL, a business must not falsely represent that goods have performance characteristics they do not have.

By implying the NPU meaningfully accelerates LLMs, when in fact it underperforms the CPU, Microsoft and Qualcomm are in breach of this section.

Claim: Goods Not of Acceptable Quality (ACL Section 54)
The Surface Pro 11 was sold as a device capable of AI acceleration. In practice, the NPU is unusable for its marketed function, meaning it is not:
- Fit for the purpose marketed
- Acceptable in performance quality

This violates the statutory guarantee of acceptable quality under ACL Section 54.

## 8. Relief Requested

I am formally requesting under Australian Consumer Law:
- An official response from Microsoft and/or Qualcomm acknowledging the performance discrepancy of the NPU in LLM inference.
- Immediate cessation of marketing that promotes AI-accelerated performance for LLMs on the Surface Pro 11 or Snapdragon X Elite, unless independently verified.
- A full refund or replacement under the Consumer Guarantees for misrepresentation of performance.
- The option to escalate this matter to the Australian Competition and Consumer Commission (ACCC) and/or local consumer protection agencies if a satisfactory remedy is not provided.

# In-Depth Benchmark Report & Legal Complaint

Llama 3.2 3B Chat on Microsoft Surface Pro 11 (Snapdragon X Elite)

## 9. Conclusion

The benchmarking results make it clear that Microsoft and Qualcomm have misled consumers about the AI capabilities of the Surface Pro 11 and Snapdragon X Elite. This misrepresentation breaches core provisions of the ACL and undermines trust in advertised AI functionality. Action must be taken to ensure performance claims reflect reality, not marketing fiction.